

# 轴承剩余使用寿命预测的 IDSA-LSTMNN<sup>\*</sup>

梁天添<sup>1</sup>, 刘 健<sup>1</sup>, 梁贺焱<sup>1</sup>, 郑皓谦<sup>1</sup>, 王 茂<sup>2</sup>

(1. 大连交通大学电气工程学院 大连, 116028)

(2. 哈尔滨工业大学空间控制与惯性技术研究中心 哈尔滨, 150001)

**摘要** 针对长短期记忆神经网络提取特征信息相关性和时间信息依赖性不足的问题, 提出基于改进双多头注意力机制的长短期记忆神经网络(improved dual stage attention-based long short-term memory neural networks, 简称 IDSA-LSTMNN), 以提高滚动轴承剩余使用寿命(remaining useful life, 简称 RUL)的预测精度。首先, 采用改进的蜘蛛蜂优化器(improved spider wasp optimizer, 简称 ISWO)优化变分模态分解(variational mode decomposition, 简称 VMD)的关键参数, 以提取更优的时频域特征; 其次, 结合时域和频域特征, 构建特征数据集; 然后, 利用非线性成分改进核主成分分析(kernel principal component analysis, 简称 KPCA)的核函数, 以优化特征降维过程; 最后, 利用多头注意力权重改进特征注意力机制和时间注意力机制提出 IDSA, 设计了轴承 RUL 预测的 IDSA-LSTMNN, 并进行了对比试验、抗噪性试验和泛化性试验。结果表明, 相较于仅具有单头注意力机制的 LSTMNN, 不同工况、不同噪声环境下, IDSA-LSTMNN 在预测精度、抗噪性和泛化性上具有显著优势。

**关键词** 轴承寿命预测; 变分模态分解; 特征降维; 长短期记忆神经网络; 改进的注意力机制

**中图分类号** TH133.33

## 引 言

轴承作为机械设备动力系统的核心部件之一, 其长期运行后会出现疲劳损伤、应力破坏等情况, 导致轴承性能退化, 影响设备的运行安全<sup>[1]</sup>。有效的轴承 RUL 预测方法有利于制定合理的检修策略, 减小维护成本, 降低故障风险, 确保机械设备安全、稳定和可靠地运行。RUL 预测方法大致分为物理模型方法和数据驱动方法<sup>[2]</sup>。物理模型方法在分析故障机制的基础上, 建立数学模型来描述机械的退化过程, 该方法存在疲劳损伤累计理论不完善、未考虑加载载荷次序对疲劳损伤的影响、模型参数众多、寿命预测耗时和预测结果精度有限等不足。

数据驱动方法利用能反映被观测对象性能退化规律的数据完成 RUL 预测, 主要包括性能退化表征指标设计与预测模型构建。目前, LSTMNN 已广泛应用于数据驱动领域。Zhu 等<sup>[3]</sup>利用双向长短期记忆神经网络(bi-directional long short-term memory neural network, 简称 BiLSTMNN), 实现了轴承 RUL 预测和健康评估。为了更好地捕捉特征信息相关性, 卷积神经网络和 BiLSTM(convolution neural network-BiLSTM, 简称 CNN-BiLSTM)与融合

注意力机制的 CNN-BiLSTM(CNN-BiLSTM with attention, 简称 CNN-BiLSTM-Att)已用于轴承 RUL 预测<sup>[4-5]</sup>。基于双注意力机制的 LSTMNN(dual stage attention mechanism based LSTMNN, 简称 DSA-LSTMNN)已用于短期光伏功率预测<sup>[6]</sup>。Li 等<sup>[7]</sup>提出多尺度卷积神经网络(multi-scale convolution neural network, 简称 MCNN)用于 RUL 预测。尽管如此, 在轴承 RUL 预测领域, 仍存在进一步提取时间信息的依赖性。

笔者提出了基于特征注意力和时间注意力的 IDSA, 以提高 LSTMNN 的预测精度。首先, 利用 ISWO-VMD 提取能表征轴承退化的时频域特征, 结合时域和频域特征构造轴承退化特征数据; 其次, 采用核函数中添加非线性成分的 KPCA 进行特征降维; 最后, 采用 IDSA-LSTMNN 预测轴承 RUL 并进行了试验验证。

## 1 轴承故障特征提取

### 1.1 时域和频域特征提取

首先, 提取的时域特征包括均值、均方根、方差、

\* 国家自然科学基金资助项目(U24B20159); 辽宁省教育厅基本科研资助项目(JYTMS20230037); 辽宁省交通科技资助项目(202243)

收稿日期: 2024-05-23; 修回日期: 2024-08-10

标准差、峭度、偏斜度、最大值以及最小值,即特征1~8。

其次,针对与轴承振动信号有关的非周期信号 $x(t)$ ,使用傅里叶变换提取多频域特征,即

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \quad (1)$$

频域特征包括均值频率、均方根频率、方差频率、标准差频率、峭度频率、偏斜度频率、最大值频率及最小值频率,即特征9~16。

## 1.2 时频域特征提取

人工选择VMD的关键参数 $K$ 和 $\alpha$ ,可能导致时频域特征提取过程具有一定的保守性。为获得更优的 $K$ 值和 $\alpha$ 值,提出ISWO-VMD算法。

蜘蛛蜂优化器(spider wasp optimizer,简称SWO)模仿蜘蛛蜂在自然界狩猎蜘蛛、追击、聚群及繁殖的过程<sup>[8]</sup>,适用于广泛的优化问题,满足不同的勘探和开发要求。但是,现有的种群初始化策略和位置参数更新容易导致SWO陷入局部最优,使得参数为局部最优解,因此提出了改进策略。

首先,使用立方映射方法初始化种群,即

$$SW_p(n+1) = 4(SW_p(n))^3 - 3SW_p(n) \quad (2)$$

其中: $SW_p(n)$ , $SW_p(n+1)$ 分别为前、后时刻种群。

其次,利用种群引导机制,增强种群个体对周围空间的探索能力,即

$$\overline{SW}_i^t = \text{rand}(\bullet)(L_i(t) + u_i(t)) - SW_i^t \quad (3)$$

其中: $\overline{SW}_i^t$ 为引导机制的解; $L_i(t)$ , $u_i(t)$ 为整个种群 $t$ 次迭代时的动态上、下界, $L_i(t) = \min(SW_i^t)$ , $u_i(t) = \max(SW_i^t)$ ;随机数 $\text{rand}(\bullet) \in [0, 1]$ ,代表动态边界的伸缩控制量。

为进一步提高算法的收敛速度与精度,在SWO位置更新中引入正余弦变换惯性权重,即

$$\omega = k \sin\left(\frac{\pi}{2} \cos\left(\sqrt{\frac{t}{t_{\max}}}\right)\right) \quad (4)$$

其中: $k \in [0, 1]$ ,为惯性权重系数; $t$ 为当前时刻迭代量; $t_{\max}$ 为最大迭代量。

引入的正余弦变换惯性权重随迭代次数变化动态调整,有助于SWO跳出局部最优,提升算法的收敛效率。引入 $\omega$ 后的位置更新过程表示为

$$\overline{SW}_i^{t+1} = \omega \overline{SW}_i^t + \mu(\overline{SW}_a^t - \overline{SW}_b^t) \quad (5)$$

其中: $\overline{SW}_a^t$ , $\overline{SW}_b^t$ 为随机位置量; $\mu$ 为搜索系数。

最后,SWO搜索采用非线性寻优策略,为防止搜索时陷入局部最优,在搜索系数中引入

$$\alpha = (1 - t/t_{\max})^2 \quad (6)$$

其中: $\alpha$ 为非线性收敛因子。

搜索系数 $\mu$ 表示为

$$\mu = \alpha |rn| r_1 \quad (7)$$

其中: $r_1 \in [0, 1]$ ;  $|rn|$ 为服从正态分布的随机数。

利用ISWO-VMD提取的时频域特征为 $\text{IMF}_1$ 能量熵和 $\text{IMF}_2$ 能量熵(即特征17和特征18)。

## 2 KPCA特征降维

经典KPCA的核函数具有一定的非线性处理能力,但经典KPCA表征轴承特征数据间较强的非线性关系比较困难。针对此问题,笔者在KPCA的核函数中添加非线性成分,则有

$$K(x_i, x_j) = \alpha e^{-\gamma \|x_i - x_j\|^2} + \beta \tan(\gamma(x_i, x_j) + r) \quad (8)$$

其中: $K(\bullet)$ 为核函数映射; $\alpha, \beta$ 为成分系数, $\alpha + \beta = 1$ ;  $x_i, x_j$ 为输入样本的特征向量;超参数 $\gamma \in [0, 1]$ ;随机数 $r \in [0, 1]$ 。

通过余弦相似度判断降维特征的相似性,即

$$S = xy / |x| |y| \quad (9)$$

其中: $x, y$ 为特征向量; $|\bullet|$ 代表向量的模。

特征处理的ISWO-VMD-KPCA步骤如下。

1) 提取时域、频域和时频域特征,其中时频域特征采用ISWO-VMD方法进行提取,具体为:①初始化ISWO的种群参数、位置参数、迭代次数和最佳适应值等;②利用种群引导机制和正余弦变换惯性权重更新蜘蛛蜂群的位置参数;③若未达到最大迭代次数,则跳转至上一步,更新ISWO的状态并进行全局寻优;否则停止寻优,输出优化后VMD的 $K$ 值和 $\alpha$ 值。

2) 对提取的时频域特征及ISWO-VMD输出的时频域特征进行贡献率筛选,得到核主成分变量。

3) 利用改进核函数的KPCA对筛选的核主成分变量进行特征降维,并判断降维特征的相似性。

## 3 IDSA-LSTMNN的RUL预测

传统LSTMNN对特征信息的相关性和时间信息的依赖性捕捉能力相对较弱,可能导致网络的精度较低。笔者提出基于特征多头注意力机制(feature multi-head attention mechanism,简称FMA)和时间多头注意力机制(time multi-head attention mechanism,简称TMA)的IDSA,进而提出IDSA-LSTMNN,从而使网络更好地捕获数据的特征和时间信息,以提升网络预测精度。

### 3.1 FMA

笔者利用FMA对LSTMNN进行输入特征提取。以时刻 $t$ 为例,对于有 $M$ 个特征的输入 $x_t = [x_{1,t}, x_{2,t}, \dots, x_{M,t}]$ ,其注意力机制权重公式为

$$e_t = \sigma(W_c x_t + b_c) \quad (10)$$

其中: $e_t = [e_{1,t}, e_{2,t}, \dots, e_{M,t}]$ ,为 $t$ 时刻输入特征的注意力机制权重系数; $W_c$ 为可训练权重矩阵; $b_c$ 为偏置向量; $\sigma(\cdot)$ 为sigmoid激活函数。

对 $e_t$ 归一化,可得第 $M$ 个注意力机制权值的特征为

$$\alpha_{M,t} = \exp(e_{M,t}) / \sum_{i=1}^M e_{i,t} \quad (11)$$

得到特征注意力机制权重 $\alpha_t = [\alpha_{1,t} \quad \alpha_{2,t} \quad \dots \quad \alpha_{M,t}]$ 后,求解加权后的特征向量 $\tilde{x}_t$ ,即

$$\tilde{x}_t = \alpha_t x_t = [\alpha_{1,t} x_{1,t} \quad \alpha_{2,t} x_{2,t} \quad \dots \quad \alpha_{M,t} x_{M,t}] \quad (12)$$

此时加权输入多头为

$$M_{\text{ulthf}} = C(h_{\text{ead},1}, h_{\text{ead},2}, \dots, h_{\text{ead},t}) \quad (13)$$

其中: $h_{\text{ead},1} = \tilde{x}_1$ ;  $h_{\text{ead},t} = \tilde{x}_t$ ;  $C$ 代表将多个加权向量进行Concat拼接。

得到FMA的表达式为

$$\hat{x}_t = M_{\text{ulthf}} + M_{\text{ulthx}} \quad (14)$$

其中: $M_{\text{ulthx}} = C(x_1, x_2, \dots, x_t)$ 。

与以往的DSA-LSTMNN<sup>[6]</sup>相比,本研究提出的FMA不仅关注原始的特征信息,还关注加权后的特征信息,动态地为模型的输入特征分配注意力机制权重,挖掘出输入特征与输出间的相关性,从而进一步提高LSTMNN的预测精度。

### 3.2 TMA

TMA为输入序列的每个历史时刻携带的时间信息分配注意力机制权重,以区分时间信息对当前时刻预测输出的影响,其独立提取不同历史时刻数据间的时间序列信息,增强了关键时刻的信息表达。

TMA的输入为LSTMNN隐藏层状态 $h_t = [h_{1,t} \quad h_{2,t} \quad \dots \quad h_{k,t}]$ 的时间序列,其中: $t$ 为模型迭代时间; $k$ 为输入序列时间窗口长度。当前时刻中每个历史时刻对应的时间注意力机制权重向量为

$$l_t = \text{ReLU}(W_d h_t + b_d) \quad (15)$$

其中: $W_d$ 为可训练的权重矩阵; $b_d$ 为偏差向量; $\text{ReLU}(\cdot)$ 为激活函数。

将每次的注意力机制权重系数归一化,得到时间的注意力机制权重。定义 $\tau$ 次的时间注意力机制权重为 $l_{\tau,t}$ ,则有 $\beta_t = [\beta_{1,t} \quad \beta_{2,t} \quad \beta_{\tau,t} \quad \dots \quad \beta_{k,t}]$ 。其中: $\beta_{\tau,t}$ 为 $\tau$ 次的注意力机制权重。

$$\beta_{\tau,t} = \exp(l_{\tau,t}) / \sum_{i=1}^k l_{i,t} \quad (16)$$

用每个对应历史时间的隐藏层状态加权,得到综合的时间序列信息状态为

$$\tilde{h}_t = \beta_t h_t = \sum_{i=1}^k \beta_{i,t} h_{i,t} \quad (17)$$

定义 $h_{\text{ead},1} = \tilde{h}_1, \dots, h_{\text{ead},t} = \tilde{h}_t$ ,则TMA的表达式为

$$h'_t = C(h_{\text{ead},1}, \dots, h_{\text{ead},t}) \quad (18)$$

综上所述,IDSA-LSTMNN算法步骤如下:

- 1) 对输入层注意力机制权重和隐藏层注意力机制权重进行参数初始化;
- 2) 由式(11)计算加权特征注意力机制权重,由式(13)进行多个特征拼接,构建FMA;
- 3) 由式(16)计算加权时间注意力权重,由式(18)进行多个时间特征拼接,构建TMA;
- 4) 更新隐藏层输出状态,若未达到最大迭代次数,则跳转至步骤3,否则停止更新时间加权特征,输出RUL预测值。

为评价所提出IDSA-LSTMNN的轴承RUL预测性能,笔者采用均方根误差(root mean square error, 简称RMSE)、平均绝对误差(mean absolute error, 简称MAE)和决定系数(coefficient of determination, 简称 $R^2$ )作为评价指标。利用皮尔逊相关系数评价特征相关性和时间依赖性,即

$$\rho_{xy} = \text{Cov}(X, Y) / \sqrt{D(X)} \sqrt{D(Y)} \quad (19)$$

其中: $X$ 为输入的特征; $Y$ 为输出真实值; $\text{Cov}(\cdot)$ 为协方差; $D(\cdot)$ 为方差。

## 4 试验结果及分析

采用IEEE PHM 2012滚动轴承开源数据,将振动传感器垂直放置于轴承外圈产生振动信号,直至轴承外圈发生故障失效。采样频率为25.6 kHz,采样间隔为10 s,采样点共计3 737 600个。选取工况2(4 200 N和1 650 r/min)下的轴承数据3(数据集2-3)进行对比试验和抗噪性试验,工况1(4 000 N和1 800 r/min)下的轴承数据3(数据集1-3)和工况3(5 000 N和1 500 r/min)下的轴承数据3(数据集3-3)进行泛化性试验,且仅考虑轴承的水平振动信号。

### 4.1 时频域特征提取和分析

VMD的模式分解个数 $K$ 和惩罚因子 $\alpha$ 对信号分解极其重要。给定 $K$ 和 $\alpha$ 的范围为: $K \in [3, 10]$ ,  $\alpha \in [100, 2 500]$ 。本研究以样本熵为适应度函数,不同优化VMD的适应度函数迭代结果如图1所示。



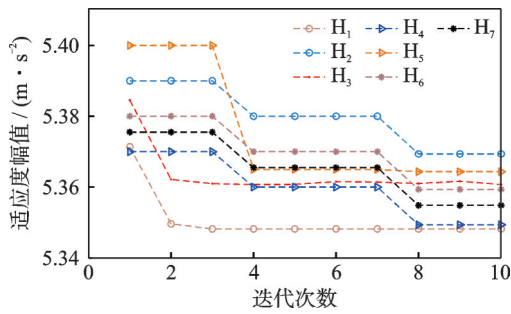


图1 不同优化VMD的适应度函数迭代结果  
Fig.1 Iterative results of fitness functions of different optimized VMDs

图1中:  $H_1 \sim H_7$ 依次代表 ISWO-VMD、改进鲸鱼优化算法<sup>[9]</sup>优化的VMD(improved whale optimization algorithm-VMD,简称IWOA-VMD)、改进灰狼优化器<sup>[10]</sup>优化的VMD(improved grey wolf optimizer-VMD,简称IGWO-VMD)、改进鸟群优化算法<sup>[11]</sup>优化的VMD(improved bird swarm algorithm-VMD,简称IBSA-VMD)、冠豪猪优化器<sup>[12]</sup>优化的VMD(crested porcupine optimizer-VMD,简称CPO-VMD)、黑翅鸢算法<sup>[13]</sup>优化的VMD(black-winged kite algorithm-VMD,简称BKA-VMD)和SWO-VMD。

由图1可知,相较于  $H_2 \sim H_7$ ,  $H_1$ 的适应度幅值更小。当迭代次数为2时,  $H_1$ 达到最优值,其余算法则迭代大于或等于8次时达到最优,表明  $H_1$ 收敛速度更快、迭代次数更少。经  $H_1$ (ISWO-VMD)优化后,得到  $K=4$ ,  $\alpha=2490$ 。  $H_1 \sim H_7$ 的信号重构结果和性能评价分别如图2和表1所示。

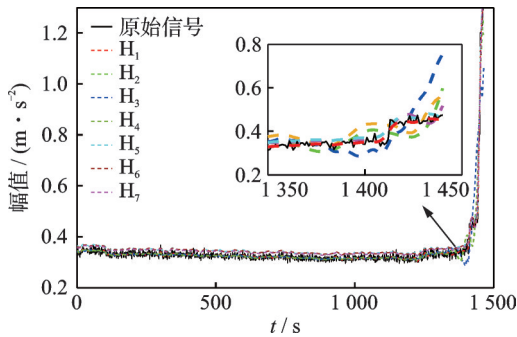


图2  $H_1 \sim H_7$ 的重构结果  
Fig.2 Reconfiguration results of  $H_1 \sim H_7$

表1  $H_1 \sim H_7$ 信号重构的性能评价

| Tab.1 Signal reconfiguration performance evaluation of $H_1 \sim H_7$ |                       |                       |         |
|-----------------------------------------------------------------------|-----------------------|-----------------------|---------|
| 模型                                                                    | RMSE                  | MAE                   | $R^2$   |
| $H_1$                                                                 | $1.26 \times 10^{-5}$ | $3.29 \times 10^{-5}$ | 0.986 7 |
| $H_2$                                                                 | $4.87 \times 10^{-5}$ | $9.07 \times 10^{-5}$ | 0.853 3 |
| $H_3$                                                                 | $5.23 \times 10^{-5}$ | $9.27 \times 10^{-5}$ | 0.803 4 |
| $H_4$                                                                 | $4.26 \times 10^{-5}$ | $8.75 \times 10^{-5}$ | 0.856 8 |
| $H_5$                                                                 | $1.76 \times 10^{-5}$ | $4.25 \times 10^{-5}$ | 0.966 7 |
| $H_6$                                                                 | $1.68 \times 10^{-5}$ | $3.75 \times 10^{-5}$ | 0.976 8 |
| $H_7$                                                                 | $2.50 \times 10^{-5}$ | $6.53 \times 10^{-5}$ | 0.953 4 |

由图2可知:在0~1 350 s,7种算法信号重构误差相差较小;在1 350~1 450 s,ISWO-VMD算法的重构信号仍能有效拟合原始信号,而其他算法的重构信号均不同程度地出现了偏差。

由表1可知:相较于  $H_2 \sim H_7$ ,  $H_1$ 的信号重构RMSE分别降低了74.12%,75.91%,70.42%,28.41%,25.00%和49.68%;MAE分别降低了63.72%,64.51%,62.40%,22.58%,12.26%和49.61%; $R^2$ 分别提高了15.63%,22.82%,15.16%,2.06%,1.01%和3.49%。

综上所述,经过种群初始化策略改进和位置参数更新后,ISWO优化了VMD的  $K$ 值和 $\alpha$ 值,从而更好地提取了信号时频域特征。

#### 4.2 特征降维的结果与分析

利用基于改进核函数的KPCA计算上述特征的各自贡献率,结果如图3所示。由图可知,特征1、特征2和特征3,即第1核主元、第2核主元及第3核主元的贡献率之和超过90%,故将上述3个特征作为核主成分变量,进行KPCA特征降维并计算特征相似性。特征降维结果如图4所示。特征相似性计算结果见表2。

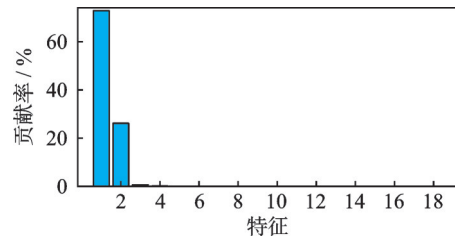


图3 特征贡献率结果  
Fig.3 Results of feature contribution rate

表2 特征相似性计算结果

| Tab.2 Feature similarity calculation results |       |       |       |
|----------------------------------------------|-------|-------|-------|
| 方法                                           | 特征1,2 | 特征1,3 | 特征2,3 |
| 经典KPCA                                       | 0.278 | 0.095 | 0.982 |
| 改进核函数KPCA                                    | 0.341 | 0.072 | 0.193 |

计算过程中,以  $J_{th}=0.5$  作为弱区分和强区分的相似度阈值( $J_{th}<0.5$ 为强区分)。由表2可知:传统KPCA降维得到的特征2和3之间的相似度为0.982,即特征2和3为弱区分,说明传统KPCA未能有效解耦特征2和3;反之,利用基于改进核函数的KPCA,降维得到每两个特征之间均为强区分,说明改进的核函数提高了KPCA对特征的解耦能力。

#### 4.3 轴承RUL预测的结果与分析

##### 4.3.1 特征相关性和时间依赖性分析

利用式(19)判断特征信息相关性和时间信息依

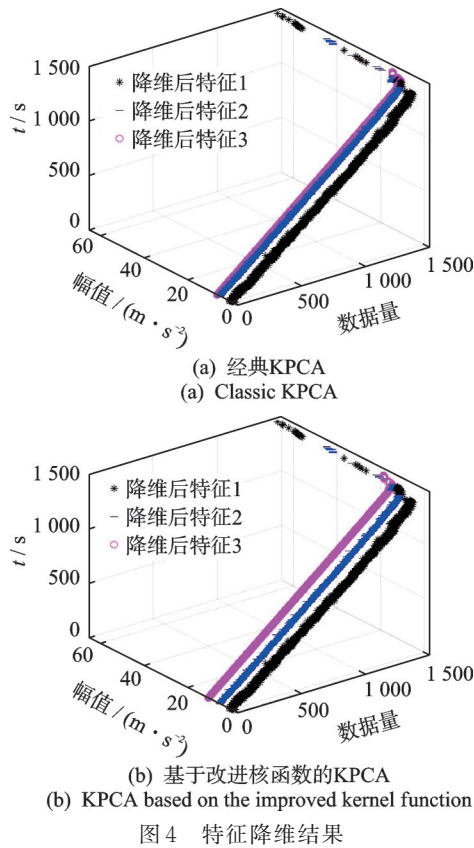


图 4 特征降维结果

赖性,并采用热力图进行分析。特征相关性和时间依赖性热力图如图 5 所示。图中: $X_1, X_2$  分别为特征信息相关性和时间信息依赖性,颜色越深,则相关性和依赖性越低;反之则越高。由图 5 可知,相较于 FMA-LSTMNN, TMA-LSTMNN 和原始 DSA-LSTMNN, IDSA-LSTMNN 的特征信息相关性分别提高了 2.61%, 91.39% 和 28.62%, 时间依赖性分别提高了 59.59%, 13.81% 和 32.43%, 验证了所提 IDSA 的有效性。

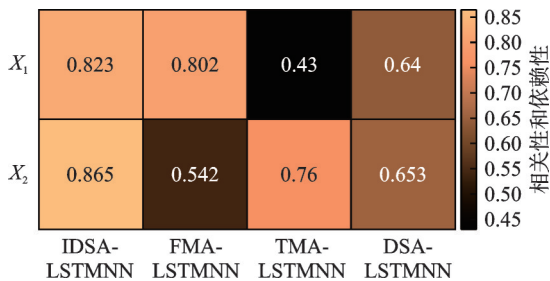


图 5 特征相关性和时间依赖性热力图

Fig.5 Heat map of the characteristic correlation and time-dependence

#### 4.3.2 对比性试验

对比性试验采用 IDSA-LSTMNN, BiLSTMNN<sup>[3]</sup>, CNN-BiLSTM<sup>[4]</sup>, CNN-BiLSTM-Att<sup>[5]</sup>, DSA-LSTMNN<sup>[6]</sup> 和 MCNN<sup>[7]</sup>。基于数据集 2-3, 不

同算法的预测结果和性能评价分别如图 6 和表 3 所示。

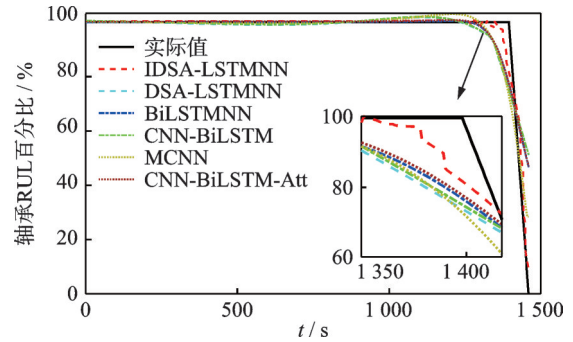


图 6 数据集 2-3 下不同算法的预测结果

Fig.6 Prediction results of different algorithms using dataset 2-3

表 3 不同算法的性能评价

| Tab.3 Performance evaluation of different algorithms |         |         |         |
|------------------------------------------------------|---------|---------|---------|
| 模型                                                   | RMSE    | MAE     | $R^2$   |
| IDSA-LSTMNN                                          | 0.854 3 | 0.270 7 | 0.989 0 |
| DSA-LSTMNN                                           | 2.464 8 | 0.705 3 | 0.944 1 |
| BiLSTMNN                                             | 2.867 0 | 0.938 1 | 0.922 4 |
| CNN-BiLSTM                                           | 1.854 3 | 0.532 4 | 0.962 7 |
| MCNN                                                 | 2.078 3 | 0.607 8 | 0.960 6 |
| CNN-BiLSTM-Att                                       | 1.423 4 | 0.372 8 | 0.964 1 |

由图 6 和表 3 可知:相较于 DSA-LSTMNN, BiLSTMNN, CNN-BiLSTM, MCNN 和 CNN-BiLSTM-Att, IDSA-LSTMNN 的预测 RMSE 分别降低了 65.34%, 70.2%, 53.92%, 58.89% 和 39.98%; MAE 分别降低了 61.62%, 71.14%, 49.51%, 55.46% 和 27.38%;  $R^2$  分别提高了 4.76%, 7.22%, 2.96%, 2.73% 和 2.58%。综上所述, IDSA-LSTMNN 的预测结果与实际值拟合度更高, 所提出的特征和时间双多头注意力机制不仅关注了特征的相关性, 还关注了时间信息的依赖性, 从而提高了网络预测性能。

#### 4.3.3 抗噪性和泛化性试验

1) 抗噪性试验。选用数据集 2-3 进行测试, 并在测试集中分别加入 SNR=2~10 dB 的白噪声, 取 5 次均值。通过控制不同变量, 获得 3 组模型。不同变量的组合设置见表 4。其中:模型  $F_1$  和  $F_2$  用于验证时间多头注意力优势;  $F_1$  和  $F_3$  用于验证特征多头注意力优势。基于不同注意力 LSTMNN 的性能评价见表 5。噪声环境下基于不同注意力 LSTMNN 的准确率见表 6。

为验证提出的 IDSA-LSTMNN 相较于常规去噪算法的有效性, 给出滑动均值滤波和小波阈值去噪方法下的预测结果。噪声环境下常规去噪方法的准确率见表 7。

表 4 不同变量的组合设置  
Tab.4 Combination settings of different variables

| 模型             | FMA | TMA |
|----------------|-----|-----|
| F <sub>1</sub> | 是   | 是   |
| F <sub>2</sub> | 是   | 否   |
| F <sub>3</sub> | 否   | 是   |

表 5 基于不同注意力 LSTMNN 的性能评价

Tab.5 Performance evaluation of different attention - based LSTMNNs

| 模型             | RMSE    | MAE     | R <sup>2</sup> |
|----------------|---------|---------|----------------|
| F <sub>1</sub> | 0.924 3 | 0.286 5 | 0.980 5        |
| F <sub>2</sub> | 2.358 3 | 0.562 4 | 0.952 3        |
| F <sub>3</sub> | 2.063 8 | 0.436 3 | 0.964 2        |

表 6 噪声环境下基于不同注意力 LSTMNN 的准确率

Tab.6 Correct rates of different attention - based LSTMNNs in the noise environment %

| 模型             | 2 dB  | 4 dB  | 6 dB  | 8 dB  | 10 dB | 无噪声   |
|----------------|-------|-------|-------|-------|-------|-------|
| F <sub>1</sub> | 97.38 | 98.02 | 98.06 | 98.07 | 98.35 | 98.45 |
| F <sub>2</sub> | 91.72 | 93.06 | 93.44 | 94.02 | 94.83 | 95.23 |
| F <sub>3</sub> | 94.26 | 95.28 | 95.52 | 96.04 | 96.25 | 96.42 |
| F <sub>4</sub> | 92.83 | 94.01 | 94.32 | 94.87 | 95.12 | 95.44 |

F<sub>1</sub>, F<sub>2</sub>, F<sub>3</sub> 和 F<sub>4</sub> 分别代表 IDSA-LSTMNN, FMA-LSTMNN, TMA-LSTMNN 和 DSA-LSTMNN

表 7 噪声环境下常规去噪方法的准确率

Tab.7 Correct rates of convention denoising methods in the noise environment %

| 模型             | 2 dB  | 4 dB  | 6 dB  | 8 dB  | 10 dB | 无噪声   |
|----------------|-------|-------|-------|-------|-------|-------|
| F <sub>5</sub> | 95.43 | 95.98 | 96.25 | 96.54 | 96.72 | 96.74 |
| F <sub>6</sub> | 97.26 | 97.53 | 97.81 | 98.01 | 98.30 | 98.32 |

F<sub>5</sub>, F<sub>6</sub> 分别为滑动均值滤波<sup>[14]</sup>和小波阈值去噪方法<sup>[15]</sup>

由表 6 可知:在无噪声和较低噪声(SNR=10 dB)的环境下,所有模型的预测准确率均较高;随

表 8 多个 LSTMNN 的性能评价

Tab.8 Performance evaluation of multi LSTMNNs

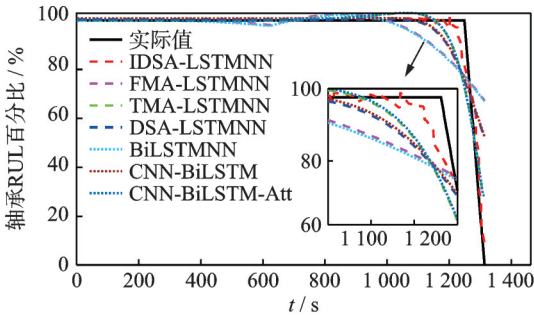
| 数据集 | 评价指标           | S <sub>1</sub> | S <sub>2</sub> | S <sub>3</sub> | S <sub>4</sub> | S <sub>5</sub> | S <sub>6</sub> | S <sub>7</sub> |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 1-3 | RMSE           | 0.810 2        | 2.753 2        | 2.088 5        | 2.494 5        | 2.078 3        | 1.862 5        | 1.523 5        |
|     | MAE            | 0.263 8        | 1.176 9        | 0.556 0        | 0.728 4        | 0.675 6        | 0.552 3        | 0.355 6        |
|     | R <sup>2</sup> | 0.986 9        | 0.920 7        | 0.967 9        | 0.946 9        | 0.943 5        | 0.955 3        | 0.964 5        |
| 3-3 | RMSE           | 0.765 3        | 2.532 6        | 2.085 3        | 3.616 1        | 2.765 3        | 2.088 3        | 1.265 3        |
|     | MAE            | 0.329 1        | 1.085 1        | 0.547 1        | 0.736 3        | 0.729 1        | 0.573 2        | 0.439 1        |
|     | R <sup>2</sup> | 0.982 6        | 0.918 8        | 0.965 9        | 0.954 4        | 0.930 1        | 0.948 8        | 0.970 1        |

S<sub>1</sub>, S<sub>2</sub>, S<sub>3</sub>, S<sub>4</sub>, S<sub>5</sub>, S<sub>6</sub> 和 S<sub>7</sub> 分别代表 IDSA-LSTMNN, FMA-LSTMNN, TMA-LSTMNN, DSA-LSTMNN, BiLSTMNN, CNN-BiLSTM 和 CNN-BiLSTM-Att

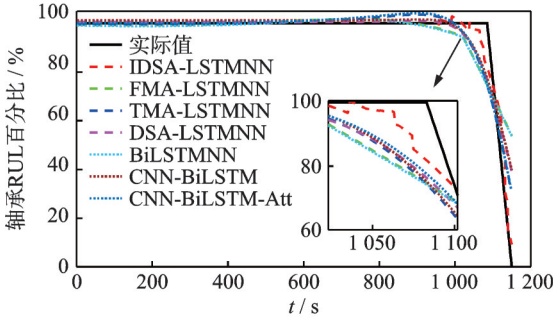
数据集 1-3 的试验结果表明:相较于 S<sub>2</sub>~S<sub>7</sub>, S<sub>1</sub> 的 RMSE 分别降低了 70.57%, 61.21%, 67.52%, 61.01%, 56.49% 和 46.81%; MAE 分别降低了 77.58%, 52.55%, 63.78%, 60.95%, 52.23% 和 25.81%; R<sup>2</sup> 分别提高了 8.36%, 1.96%, 4.22%,

着噪声的不断加强, F<sub>2</sub>, F<sub>3</sub> 和 F<sub>4</sub> 的预测准确率下降较为明显, 而 F<sub>1</sub> 在不同噪声环境下的预测准确率均超过 97%。由表 7 可知:相较于常规去噪方法, F<sub>1</sub>, F<sub>5</sub> 和 F<sub>6</sub> 的准确率也略有提高。综上所述, 在轴承 RUL 预测领域, 所提出的 IDSA-LSTMNN 不仅精度更高, 且抗噪性更好。

2) 泛化性试验。以数据集 1-3 和数据集 3-3 作为测试集, 进一步验证本研究方法相较于其他 LSTMNN 的泛化性。多个 LSTMNN 预测结果和性能评价分别如图 7 和表 8 所示。



(a) 数据集 1-3 下多个 LSTMNN 的预测结果  
(a) Prediction results of multi LSTMNNs using dataset 1-3



(b) 数据集 3-3 下多个 LSTMNN 的预测结果  
(b) Prediction results of multi LSTMNNs using dataset 3-3

图 7 多个 LSTMNN 的预测结果

Fig.7 Prediction results of multi LSTMNNs

4.59%, 3.31% 和 2.32%。数据集 3-3 的试验结果表明:相较于 S<sub>2</sub>~S<sub>7</sub>, S<sub>1</sub> 的 RMSE 分别降低了 69.78%, 63.3%, 78.84%, 72.32%, 63.35% 和 39.51%; MAE 分别降低了 69.67%, 39.84%, 55.3%, 54.86%, 42.58% 和 25.05%; R<sup>2</sup> 分别提高了 6.94%, 1.73%,



2.95%,5.64%,3.56%和1.28%。综上所述,提出的 IDSA-LSTMNN 在不同工况下的预测性能均优于其他模型。

为了进一步验证本研究方法的泛化性,选取美国辛辛那提大学全寿命轴承数据,在 SNR=8 dB 环境中,测试多个 LSTMNN 的预测准确率及各模型的计算时间如表 9 所示。

表 9 多个 LSTMNN 的预测准确率和计算时间  
Tab.9 Prediction correct rates and calculation times of multi LSTMNNs

| 模型             | 准确率/% | 计算时间/s |
|----------------|-------|--------|
| S <sub>1</sub> | 98.26 | 88.2   |
| S <sub>2</sub> | 94.27 | 83.3   |
| S <sub>3</sub> | 96.53 | 85.2   |
| S <sub>4</sub> | 94.84 | 86.4   |
| S <sub>5</sub> | 93.42 | 87.5   |
| S <sub>6</sub> | 95.23 | 92.6   |
| S <sub>7</sub> | 97.26 | 94.2   |

由表 9 可知:在较强噪声环境下,所提 S<sub>1</sub> 模型预测准确率达到 98.26%,远高于其他模型;相较于原始模型 S<sub>4</sub>,S<sub>1</sub> 尽管添加了多头注意力机制,运行时间仅增加了 1.6 s,准确率提升了 3.61%。这表明本研究提出的模型在增加较少计算成本的同时,获得了更高的轴承 RUL 预测精度。

4.3.4 自适应工况下的噪声测试

轴承实际工作时不仅受到噪声干扰,也受到工况变化的影响,且较难获得足够的故障工况数据。因此,本研究进行了自适应工况下的噪声测试。在测试阶段,模拟-3~3 dB 噪声环境,选用轴承转速为 1 650 r/min 工况下的数据训练模型,转速为 1 500 和 1 800 r/min 工况下的数据进行测试。基于上述方式,每分贝噪声下,多个 LSTMNN 在 1 种工况训练、其余工况测试的平均准确率及其在不同噪声下的准确率分别如图 8 和表 10 所示。

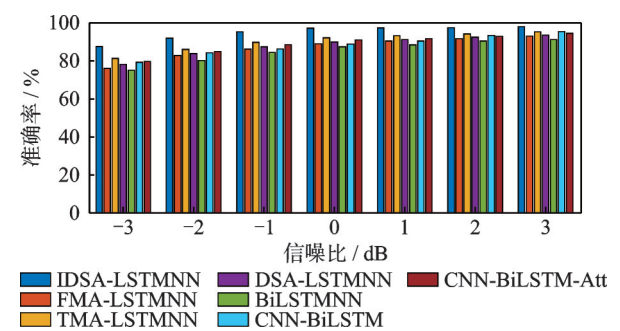


图 8 多个 LSTMNN 在不同噪声下的准确率  
Fig.8 The correct rates of multi LSTMNNs under different noise

表 10 1 种工况训练、其余工况测试的平均准确率  
Tab.10 Average correct rates of one working condition trains and the other working conditions test

| 噪声/<br>dB | 模型             |                |                |                |                |                |                |
|-----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|           | S <sub>1</sub> | S <sub>2</sub> | S <sub>3</sub> | S <sub>4</sub> | S <sub>5</sub> | S <sub>6</sub> | S <sub>7</sub> |
| -3        | 87.61          | 76.10          | 81.27          | 78.13          | 75.13          | 79.25          | 79.70          |
| -2        | 92.05          | 82.83          | 86.03          | 83.87          | 80.27          | 84.27          | 85.95          |
| -1        | 95.24          | 86.14          | 89.75          | 87.37          | 84.37          | 86.33          | 88.46          |
| 0         | 97.14          | 88.94          | 92.15          | 89.97          | 87.47          | 88.76          | 92.06          |
| 1         | 97.26          | 90.33          | 93.18          | 91.18          | 88.34          | 90.34          | 92.64          |
| 2         | 97.48          | 91.72          | 94.20          | 92.38          | 90.38          | 93.36          | 92.87          |
| 3         | 98.01          | 93.01          | 95.28          | 93.56          | 91.25          | 95.35          | 94.56          |
| 平均值       | 94.97          | 87.01          | 90.27          | 88.07          | 85.31          | 88.23          | 89.46          |

由表 10 和图 8 可知:S<sub>1</sub>(IDSA-LSTMNN)在-3~3 dB 噪声环境中,每分贝噪声下的预测准确率均高于其他 6 种模型,且相较于 S<sub>2</sub>~S<sub>7</sub>,S<sub>1</sub> 的平均准确率分别提高了 9.14%,5.21%,7.83%,11.32%,7.63%和 6.15%。综合分析,所提出的 IDSA-LSTMNN 因使用了特征多头注意力和时间多头注意力机制,模型兼顾特征相关性和时间依赖性,即使在强噪声环境下,也具有较高的轴承 RUL 预测精度。

5 结 论

1) 利用立方映射、种群引导机制和正余弦惯性权重方法改进了蜘蛛蜂优化算法,寻优了 VMD 的模态分解个数 K 和惩戒因子  $\alpha$ ,并提取了能量熵作为特征向量。试验结果表明,ISWO-VMD 对特征提取的准确率达到 98.67%,超过了其他方法的准确率。

2) 利用非线性成分改进了 KPCA 的核函数,并通过余弦相似度来判断特征相似性。计算结果表明,基于改进核函数的 KPCA 有效降低了特征间的相似度,提高了对特征的解耦能力。

3) 将特征多头注意力机制和时间多头注意力机制混合使用,得到 IDSA,并提出 IDSA-LSTMNN 预测模型。同一工况、同一噪声环境下的对比试验表明,相较于其他方法,IDSA-LSTMNN 的剩余使用寿命预测精度有明显提高;同一工况、不同噪声环境下的抗噪性试验表明,所提出模型有较高的抗噪性;不同工况、同一噪声环境下的泛化试验表明,IDSA-LSTMNN 对于不同的工况均有较好的适应性;不同工况、不同噪声下的自适应工况试验表明,在 SNR=-3~3 dB 环境中,模型的平均准确率为 94.97%,模型准确度和抗噪性能均较好。

## 参 考 文 献

- [1] 郭晓静, 负玉晶, 徐晓慧. 基于深度学习方法的航空发动机寿命预测模型[J]. 振动、测试与诊断, 2024, 44(2): 330-336.
- GUO Xiaojing, YUN Yujing, XU Xiaohui. Prediction model of aero-engine remaining useful life based on deep learning method[J]. Journal of Vibration, Measurement & Diagnosis, 2024, 44(2): 330-336. (in Chinese)
- [2] 王前, 黄强, 江星星, 等. 基于融合特征与数模联动的轴承寿命预测方法[J]. 振动、测试与诊断, 2023, 43(4): 705-711.
- WANG Qian, HUANG Qiang, JIANG Xingxing, et al. Research on remaining useful life prediction of rolling bearings based on fusion feature and model-data-fusion[J]. Journal of Vibration, Measurement & Diagnosis, 2023, 43(4): 705-711. (in Chinese)
- [3] ZHU Q X, ZHOU Z, LI Y, et al. Contrastive BiLSTM-enabled health representation learning for remaining useful life prediction[J]. Reliability Engineering & System Safety, 2024, 249: 110210.
- [4] GUO J Y, WANG J, WANG Z, et al. A CNN-BiLSTM bootstrap integrated method for remaining useful life prediction of rolling bearings[J]. Quality and Reliability Engineering International, 2023, 39(5): 1796-1813.
- [5] ZHANG W, JIN S, BIAN G, et al. A method for sound speed profile prediction based on CNN-BiLSTM-Attention network[J]. Journal of Marine Science and Engineering, 2024, 12(3): 414.
- [6] 杨晶显, 张帅, 刘继春, 等. 基于VMD和双重注意力机制LSTM的短期光伏功率预测[J]. 电力系统自动化, 2021, 45(3): 174-182.
- YANG Jingxian, ZHANG Shuai, LIU Jichun, et al. Short term photovoltaic power prediction based on variational mode decomposition and long short-term memory with dual-stage attention mechanism LSTM[J]. Automation of Electric Power Systems, 2021, 45(3): 174-182. (in Chinese)
- [7] LI H, ZHAO W, ZHANG Y, et al. Remaining useful life prediction using multi-scale deep convolutional neural network[J]. Applied Soft Computing, 2020, 89: 106113.
- [8] ABODEL-BASSET M, MOHAMED R, JAMEEL M, et al. Spider wasp optimizer: a novel meta-heuristic optimization algorithm[J]. Artificial Intelligence Review, 2023, 56(10): 11675-11738.
- [9] 王伟平, 钟锐, 王青山. 基于改进鲸鱼算法的复合材料层合板优化设计[J]. 振动与冲击, 2023, 42(19): 294-300.
- WANG Weiping, ZHONG Rui, WANG Qingshan. Optimization design of composite laminate based on improved whale optimization algorithm[J]. Journal of Vibration and Shock, 2023, 42(19): 294-300. (in Chinese)
- [10] 薛阳, 燕宇铖, 贾巍, 等. 基于改进灰狼算法优化长期记忆网络的光伏功率预测[J]. 太阳能学报, 2023, 44(7): 207-213.
- XUE Yang, YAN Yucheng, JIA Wei, et al. Photovoltaic power prediction model based on IGWO-LSTM[J]. Acta Energeticae Solaris Sinica, 2023, 44(7): 207-213. (in Chinese)
- [11] CHEN X. Vehicle feature recognition via a convolutional neural network with an improved bird swarm algorithm[J]. Journal of Internet Technology, 2023, 24(2): 421-432.
- [12] ABDEL-BASSET M, MOHAMED R, ABOUHAUWASH M. Crested porcupine optimizer: a new nature-inspired metaheuristic[J]. Knowledge-Based Systems, 2024, 284: 111257.
- [13] WANG J, WANG W, HU X, et al. Black-winged kite algorithm: a nature-inspired meta-heuristic for solving benchmark functions and engineering problems[J]. Artificial Intelligence Review, 2024, 57(4): 98.
- [14] 田博文, 张志禹, 杨梦飞. 基于多次滑动均值滤波的混合储能功率分配与定容研究[J]. 电工技术学报, 2024, 39(5): 1548-1564.
- TIAN Bowen, ZHANG Zhiyu, YANG Mengfei. Research on hybrid energy storage power allocation and capacity determination based on multiple moving average filtering[J]. Transactions of China Electrotechnical Society, 2024, 39(5): 1548-1564. (in Chinese)
- [15] 胡涛, 申立群, 雷鹏, 等. 基于惯性传感器的输电线路舞动信号处理技术[J]. 振动与冲击, 2023, 42(19): 49-57.
- HU Tao, SHEN Liqun, LEI Peng, et al. Transmission line galloping signal processing technology based on inertial sensors[J]. Journal of Vibration and Shock, 2023, 42(19): 49-57. (in Chinese)



**第一作者简介:**梁天添,男,1986年11月生,博士、副教授。主要研究方向为高速列车故障诊断、轴承故障诊断等。曾发表《Finite frequency fault estimation and fault-tolerant control for dynamics of high-speed train based on descriptor systems》(《Transactions of the Institute of Measurement and Control》2023, Vol.45, No.2)等论文。

E-mail: liangtiantian1122@163.com